Subordination strategies in South America: Nominalization

Van Gijn, Rik

Abstract: This chapter argues that nominalization, as a subordination strategy, is significantly more pervasive in South America than would be predicted on the basis of global patterns. The patterns found within South America are most consistent with a scenario of several smaller spreads, possibly promoted by a few language families with major extensions (e.g. Quechuan, Tupian, Cariban).

12 Subordination strategies in South America: nominalization

Rik van Gijn

This chapter argues that nominalization, as a subordination strategy, is significantly more pervasive in South America than would be predicted on the basis of global patterns. The patterns found within South America are most consistent with a scenario of several smaller spreads, possibly promoted by a few language families with major extensions (e.g. Quechuan, Tupian, Cariban).

1 Introduction

Nominalized subordinate clauses are extremely common in South American languages and have been mentioned repeatedly as an areal or regional feature for geographic zones of different extensions. For example Dixon and Aikhenvald (1999: 9) claim for the vast Amazon basin: "Subordinate clauses typically involve nominalized verbs, with the type of subordination marked on the verb," and Crevels and Van der Voort (2008) mention "subordination through nominalization" as one of the areal features of the Guaporé-Mamoré area in northeast Bolivia and Rondônia in western Brazil. But nominalized subordinate clauses are also common in the Andean linguistic area: Mapudungun and the Quechuan and Aymaran languages (see e.g. Torero 2002, Adelaar with Muysken 2004) all have several types of nominalized clauses.

Such areal claims suggest a scenario of diffusion through contact of this structure rather than through inheritance or due to chance. Although it is probably not possible to prove or disprove the contact-induced diffusion of nominalized clauses beyond doubt, we can evaluate its likelihood against two other possible explanations: genealogical inheritance and chance. I attempt to do this by answering the following two questions.

The larger project that this study is part of has been made possible by the support of Netherlands Organization for Scientific Research (NWO – grant 275–89–006), carried out at the Radboud University Nijmegen. This support is gratefully acknowledged. I furthermore thank the editors for useful comments on earlier versions of this chapter. Remaining errors are mine.

(i) *Is the distribution of nominalized subordinate clauses geographically skewed towards South America?*

A key element for a claim of contact-induced areal spread of a feature is that its distribution should be geographically skewed, i.e. present or preferably even abundant in certain geographic zones while scarce or absent in others (especially adjacent ones). The question of geographic skewing can be answered by looking at the distribution of nominalized subordinate clauses on a global scale, based on the study by Cristofaro (2003). If the presence of nominalized structures does not differ significantly from global distribution, chance, or some more general (e.g. cognitive or diachronic) principle may better explain the presence of the nominalized structures. We can also address this question by looking at South American languages only: do certain geographic zones like the Amazon basin, the Andes, and the Guaporé-Mamoré area stand out from other areas with respect to this feature? If so, contact may still be the factor with the greatest explanatory power, even if the general distribution in South America does not differ significantly from that of global samples.

(ii) *Is there variation within the group of nominalized structures, and is that geographically skewed?*

"Nominalized subordinate clauses" is a very general term that potentially encompasses a host of different structures. Typological research (e.g. Comrie 1976; Koptjevskaia-Tamm 1993; Malchukov 2006; Comrie and Thompson 2007) has shown that nominalized constructions can differ from each other on various parameters, both in terms of their morphosyntax and in their semantics. Therefore, before we can claim diffusion of nominalized subordinate clauses, we need to make sure that we are comparing like with like. If there is much internal variation within the group of nominalized structures, it might shed a more differentiated light on the inheritance or diffusion through contact of particular nominalized structures. It might, for instance, differentiate Andean nominalized structures from Amazonian ones, or it might perhaps show that Quechuan nominalized structures are special structurally, a factor best explained in terms of genealogical inheritance.

The chapter is set up as follows. In Section 2 I will discuss some preliminaries, including the definition of nominalization as a subordination strategy, the description of the South American sample used in this study, and the way in which I measure distances between constructions. Section 3 addresses question (i) above, by comparing the South American sample to the global sample used in Cristofaro (2003) as well as by looking at the South American sample itself. Section 4 discusses the nominalized structures found in South America in more detail, and discusses the internal variation found between them (question (ii) above). Section 5, finally, is a discussion of the results, in which I evaluate possible explanations for the distributional patterns found in South American nominalized clauses.

Table 12.1 *Semantic relations considered for subordination strategies*

Complement relations	phasal (start/finish), modal-ability (can, be able, know how to), desiderative (want), direct manipulation (make, cause), knowledge (know that), perception (see, hear), indirect utterance (say that), evaluative constructions (that 'X' is good/bad)
Relative relations	S relatives, A relatives, O relatives
Adverbial relations	temporal (simultaneous, successive), reason, location, purpose (motion, other), condition (hypothetical, counterfactual)

2 Preliminaries

The data presented in the present chapter are part of a larger project on subordination strategies. The definition of nominalized clauses is based on the set-up of this larger project, so it is useful to start this section by briefly outlining the bigger project.[1] The project "Subordination strategies in South American languages" aims at comparing morphosyntactic strategies that languages employ to encode certain semantic relations between events, and measuring the distance between these strategies. The semantic relations taken into account are given in Table 12.1.

These semantic relation types are in large part based on Cristofaro (2003), which makes a comparison with her results feasible. Moreover, as argued by Cristofaro (2003), they form a collection of semantic relations that have different basic semantic parameters, so they are likely to yield most if not all subordination strategies in a language. Based on these semantic relation types, different constructions in each of the languages that encode them are selected for comparison. These constructions may in principle differ widely from each other, from bi-clausal structures to derivational affixing, and from fully finite structures to bare infinitives and nominalized structures. In order to be able to compare all these different structures to each other, a questionnaire was developed which targets the subatoms (individual morphosyntactic characteristics) of the constructions. Questions fall into five thematic realms: finiteness, nominalization, flagging, integration, and linearization.

Finiteness relates to the verbal categories that can be marked on the dependent unit (as opposed to an independently used verb) and also pertain to it.[2] Since languages can differ considerably in terms of the categories they can

[1] For a more detailed description, see Van Gijn and Hammarström, in prep.

[2] This is meant to exclude instances of clitics that happen to be placed on the dependent unit, but have scope over the whole sentence.

mark on a verb, I focus on the more common ones: subject agreement, object agreement, tense, aspect, event modality, epistemic modality, and evidentiality. *Nominalization* relates to the nominal categories that can be marked on the dependent unit, also focusing on the more common categories: can they take case/adpositions, can they combine with determiners, can their subject or object be encoded as a possessor, can they trigger agreement on other elements, and finally can they take nominal plural markers? *Flagging* targets overt linguistic signs of dependency, such as complementizers, subordinators, dependency markers, but also nominalizing affixes and special (i.e. deviant from main clauses) markers for tense aspect and/or modality. *Integration* concerns (apart from whether verbal categories can be marked independently for the dependent unit) whether the independent unit can be negated separately, contiguity of the main and dependent units, or even morphological fusion. *Linearization*, finally, looks at the position of dependency markers with respect to the dependent unit, the position of the dependent unit with respect to the main unit, and, specifically for relative relations, the position of the relativized noun with respect to the relative clause (or functional equivalent).

Constructions, including nominalizations, can differ from or be similar to each other in all of these respects. This gives a fine-grained comparative measure on the construction level, and it can also produce a measure on the language level, which involves a number of technicalities that need not concern us for this chapter (see Van Gijn and Hammarström in prep. for more details), since we will only measure distances between constructions.

Nominalizations can now quite straightforwardly be defined in terms of the questions on nominalization mentioned above in this chapter. Since nominalization forms the heart of the chapter, I will zoom in on the questions concerning nominalization in slightly more detail. The questions and their possible answers are given in Table 12.2.

The dependent EDU (event-denoting unit) is the element that refers to the event that either modifies another event (adverbial relations), modifies a participant in another event (relative relations), or is entailed by another event (complement relations) – see Cristofaro (2003). The nominal characteristics of a dependent EDU that are coded in the questionnaire are the ability to be case-marked or to combine with an adposition, the potential to be modified by a determiner or an attributively used demonstrative, the possibility of encoding either the subject and/or the object of the dependent EDU as a possessor, the potential to trigger agreement on other elements (e.g. subject agreement, object agreement, adjective agreement), and finally whether the dependent EDU can be specified for nominal number. As "subject" I mean the A participant in transitive clauses, plus the S participant in intransitive clauses insofar as the latter is encoded in the same way as the former. With "object" I mean the P (or O) participant in a transitive clause, plus the S participant in an intransitive

Table 12.2 *Questions on nominalization*

Question	Answer key
Can the dependent EDU be case-marked or marked by an adposition?	Y/N
Can the dependent EDU take determiners or attributive demonstratives?	Y/N
Can the subject of the dependent EDU be marked as a possessor?	A=no, B=yes, C=possessor undistinguishable from verbal marking
Can the object of the dependent EDU be marked as a possessor?	A=no, B=yes, C=possessor undistinguishable from verbal marking
Can the dependent EDU trigger agreement on other elements?	Y/N
Can the dependent EDU be marked for nominal number?	Y/N

clause insofar as the latter is encoded in the same way as the former. Many South American languages code the possessor in the same way as one of the core arguments, so that it becomes impossible to tell whether the argument encoding in a dependent clause marks a possessor or a subject/object. For this situation, a third possible answer has been created in the questionnaire. For the possessor questions there is a third possible answer which is relevant for many South American languages (see e.g. Dixon and Aikhenvald 1999: 9), namely that the encoding of possessor is identical to the encoding of subject or object in independent clauses.

A nominalization can now be defined as a construction for which the answer to one of the questions in Table 12.2 is "yes" (or "B" in the case of the possessor questions – isomorphic possessors are not counted). This means that nominalization is defined independently from deverbalization, as well as from syntactic function.

The sample used for this chapter consists of forty languages spoken throughout South America, but with a clear focus on western South America, where language diversity is greatest.

3 The distribution of nominalization as a subordination strategy in South America vs. the world

Nominalization is a very widespread subordination strategy in South American languages. It has been mentioned as an areal feature for larger and smaller regions (e.g. Dixon and Aikhenvald 1999 for the Amazon, Crevels and Van

der Voort 2008 for the Guaporé-Mamoré), but the distribution seems to extend well beyond both. The question that I address in this section is whether this distribution stands out in some way compared to the distributional patterns of nominalized subordinate clauses on a global scale.

From more theoretical and diachronic perspectives on language, it seems unsurprising that nominalization is a prominent subordination strategy. For instance, Heine and Kuteva (2007) discuss diachronic pathways through which subordinate clauses may arise. For both complement clauses and adverbial clauses, they propose two main pathways, expansion and integration. The latter refers to the reinterpretation of two separate clauses as a single, complex clause; the former is meant as the reinterpretation of a noun phrase as a clause, which is mentioned as an important pathway in particular for complement clauses and adverbial clauses (see Deutscher 2009 for relative clauses). Crucially for this paper, Heine and Kuteva (2007: 216–217) mention that "nominal" characteristics often survive such a process of reinterpretation:

a. The marker of subordination resembles a grammatical form associated with noun phrase structure, such as a marker of case, gender, or definiteness, or an adposition.
b. The verb of the subordinate clause is frequently non-finite, encoded like an infinitival, gerundival, participial, or nominalized constituent and takes the case marking of a corresponding nominal participant.
c. The arguments of the subordinate clause are encoded in a form that tends to differ from that of the main clause.
d. The agent or notional subject takes a genitive/possessive or other case form, typically having the appearance of a genitival modifier of the subordinate verb.
e. The patient or notional object may also take a genitive/possessive or other case form.
f. There are severe restrictions on distinctions such as tense, aspect, modality, negation, etc. that can be expressed – in fact, such distinctions may be absent altogether.

Heine and Kuteva's surviving nominal traits *a–e* correspond to some of the questions on nominalization in the questionnaire discussed above. Since characteristic *f* is treated as separate from nominalization, it does not play a role in defining a nominalized construction, even though the degree of lack of verbal features does play a role in measuring distances between constructions.

There is, moreover, a functional motivation for a connection between subordinate clauses and nominalization, as discussed by Malchukov (2006), based on Croft (1991): on the one hand, subordinate clauses express events, which are normally expressed by verbs; they have a time reference and possibly an internal temporal structure, and they have participants in verbal semantic roles like agent and patient. On the other hand, they function as arguments of verbs, or possibly of adpositions (e.g. to form adverbial modifications). There is, in other words, a category-mismatch between the lexical root (verbal) and the argument function (which expects a referential expression).

Cristofaro (2003) goes one step further by suggesting a deeper, cognitive explanation for the predominance of nominalized structures for referring to dependent events. She argues that dependent events are processed differently than independent events: "By virtue of lacking an autonomous profile, dependent SoAs [States of Affairs – RG] are not scanned sequentially, but construed as a unitary whole, just like things" (p. 262). This, in Cristofaro's view, may explain the predominance of nominal categories in dependent clauses, although nominal characteristics are not equally likely for every relation type, a point to which I come back to below.

In other words, there may be independent reasons that nominalization predominates as a subordination strategy, reasons quite separate from contact or inheritance. It is therefore useful to compare the distribution of data found in the South American sample with distributions on a global scale, provided by Cristofaro (2003). Cristofaro's study contains information on several semantic relations between events in a typologically balanced sample of 80 languages. As mentioned above, many of the semantic relations in her study are taken as a basis in the South American study. Since Cristofaro also looks at constructions for which she keeps track of verbal and nominal categories that can be found on dependent EDUs, this makes the two studies comparable to a large degree. The nominal categories tracked by Cristofaro are case marking and possession. We can now look at two points of comparison between the South American data and the global data: the number of nominalizations found and the distribution of these nominalizations over the different semantic relation types. In order to make the results maximally comparable, I only look at those relation types that are present in both studies, and nominalization will be defined only on the basis of case marking and possession.

Cristofaro (pp. 311–333) lists a total of 423 constructions that are taken into consideration. Sixteen constructions had to be discounted for the comparison because the semantics of those constructions were not, or not sufficiently, comparable,[3] represented in the South America questionnaire, leaving a total of 407 constructions. About a quarter of these constructions can be classified as nominalized.[4] At the language level, a little less than half (38) of the languages in her sample have nominalized constructions. In my sample of South American languages, fewer than 40 percent of the total number of

[3] I have not counted constructions in Cristofaro's appendix that encoded "order" manipulation, "before" relations, and utterance. Although the last is also a category in the South America database, I have only coded indirect utterance constructions. Cristofaro also counts direct utterance if it is the only way to encode speech complements. I consider Cristofaro's "propositional attitude" relation type comparable to my "evaluative" relation type, because they have the same semantic outline.

[4] The actual number may be higher, since only case marking and possession are taken into account, so this should be taken as a minimum number.

Table 12.3 *Comparison of global and South American distributions of nominalized structures*

	Cristofaro	Van Gijn
languages	80	40
constructions	407	230
nominalized constructions	101 (24.8%)	91 (39.6%)
languages with nominalized construction(s)	38 (47.5%)	36 (90%)

constructions are nominalized,[5] and almost all languages (90 percent) have at least one nominalized construction, as shown in Table 12.3.

The proportion of South American languages that have nominalized constructions compared to the global sample is highly significant ($p = 3.26e\text{-}06$ in a Fisher's exact test), as is the number of nominalized constructions as a proportion of the total number of constructions ($p = 0.0001$). On the first count then, nominalized structures appear in significantly higher numbers in South America than would be expected on the basis of the global patterns, both in terms of number of constructions and in the number of languages that have nominalized constructions.

A second comparison that can be made to Cristofaro's study is the distribution of nominalized structures over different semantic relation types. Cristofaro (p. 263) mentions that nominalized structures (defined as having case marking possibilities)[6] are not evenly distributed over the semantic types, but rather follow a hierarchy, given in (1):

(1) Case/adposition hierarchy (slightly adapted from Cristofaro 2003: 230)
Modals, Phasals, Purpose, Desideratives, Manipulatives, Perception, Temporal, Reason > A/S/O relativization > Condition, Knowledge, Utterance, Propositional attitude.

The hierarchy should be read as follows: if a nominalized form (i.e. one that can take case/adposition marking) is used to encode the dependent event at a point on the hierarchy, the points to its left will also allow a nominalized form.

The specific distribution per type in Cristofaro's and my samples are compared in Table 12.4. The numbers do not add up to reflect the number of constructions, because there is often a one-to-many relationship between constructions and meaning. The number of nominalized manipulation predicates is low in Cristofaro's distribution in part because I have only counted direct (make) manipulation.

[5] I have discounted utterance and location clauses, since Cristofaro does not consider the latter; for the former, see above.

[6] No hierarchy was proposed for possession, but it follows a similar pattern (Cristofaro 2003: 235).

Table 12.4 *Comparison of distribution of nominalized structures per semantic relation type*

	Cristofaro	%	Van Gijn	%	p-value
modals	16	7.5	16	5.6	0.46
phasals	16	7.5	16	5.6	0.46
purpose	29	13.6	29	10.1	0.26
desideratives	19	8.9	26	9.1	1
manipulatives	5	2.3	2	0.7	0.14
perception	21	9.9	22	7.7	0.42
temporal	24	11.3	36	12.6	0.76
reason	24	11.3	25	8.7	0.36
relativization	25	11.7	43	15.0	0.36
condition	10	4.7	21	7.3	0.26
knowledge	10	4.7	21	7.3	0.26
utterance	9	4.2	12	4.2	1
propositional attitude	5	2.3	18	6.2	0.05
total	213		287		

Although there are differences between the distributions in the samples, none of them is significant, which means that, if there were to be evidence of contact-induced diffusion of nominalized constructions, this is not connected to a particular semantic field. So a next step we can take is to look in more detail into the structural properties of the nominalizations of South American languages.

4 Types of nominalizations in South America and their distribution

It has been recognized by many scholars that the typology of nominalization shows quite a bit of internal structural variation cross-linguistically. For instance, nominalizations may differ from each other in how they encode core arguments (see e.g. Koptjevskaja-Tamm 1993), or in the extent to which they allow for verbal and nominal categories to be marked on the nominalized verb (see e.g. Malchukov 2006), nominalizations may be flagged in different ways, and of course they can differ in which semantic relation types they can encode.

Because of the potential variation within the group of nominalized structures, it makes sense to evaluate the homogeneity of the nominalizations across the continent, and to see whether the internal variation found can best be explained as a geographic (contact) signal or a phylogenetic signal. For a first impression of the internal variation of nominalized constructions, see Figure 12.1, which gives a visual representation in the form of a NeighborNet network (Bryant and Moulton 2004) of the distance between the constructions

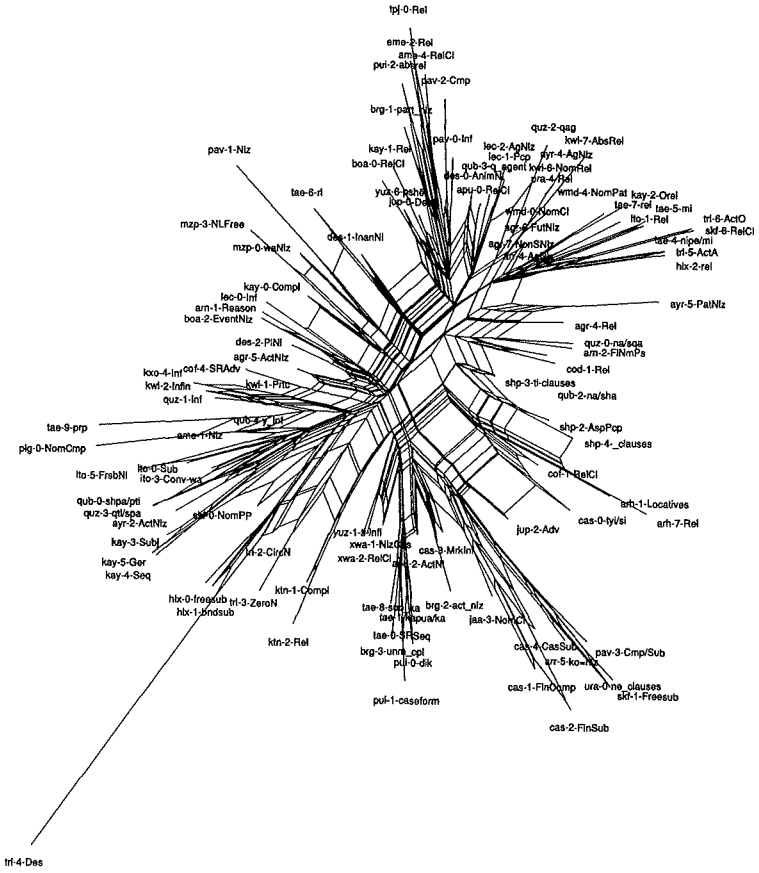


Figure 12.1 NeighborNet of nominalizations as subordination strategies in the languages of the sample

based on similarity among the input features. The sheer number of the constructions renders the figure rather difficult to read, but the star-shaped form and the general lack of tree-like branches indicate that the nominalizations that are used as subordination strategies are far from homogeneous.

In the remainder of this section, I will look in greater detail at the different nominalizing subordination strategies found in South American languages. I take semantics as a basis for comparison, based on the assumption that, if a language borrows a construction, or if two constructions in different languages converge as a result of contact, they will most likely have comparable semantics.

Table 12.5 *Overlap of semantic relation types*

	TEMP	REAS	LOC	PURP	COND	PHAS	MODL	DESID	MANIP	PERC	KNOW	IND UTT	EVAL	S-REL	A-REL	O-REL
TEMP		41	13	31	41	5	10	15	2	18	13	11	13	10	10	13
REAS	41		11	23	33	4	7	10	2	10	12	8	10	8	8	12
LOC	13	11		7	9	7	6	8	0	9	10	5	10	10	10	12
PURP	31	23	7		21	7	10	17	3	15	12	8	10	9	9	10
COND	41	33	9	21		2	5	7	0	7	8	6	7	6	6	8
PHAS	5	4	7	7	2		25	30	12	11	15	7	10	1	1	4
MODL	10	7	6	10	5	25		33	16	15	16	10	12	2	2	4
DESID	15	10	8	17	7	30	33		23	22	26	15	20	6	6	10
MANIP	2	2	0	3	0	12	16	23		4	4	2	3	0	0	0
PERC	18	10	9	15	7	11	15	22	4		24	13	16	8	8	7
KNOW	13	12	10	12	8	15	16	26	4	24		20	14	9	9	14
IND UTT	11	8	5	8	6	7	10	15	2	13	20		10	4	4	8
EVAL	15	10	10	10	7	10	12	20	3	16	14	10		4	4	6
S-REL	10	8	10	9	6	1	2	6	0	8	9	4	4		43	37
A-REL	10	8	10	9	6	1	2	6	0	8	9	4	4	43		35
O-REL	13	12	12	10	8	4	4	10	0	7	14	8	6	37	35	

However, defining comparable semantics can be a complex task, since we do not know which semantic building blocks of the different relation types are relevant. The way I approach this problem is to look at every semantic relation type defined in the questionnaire separately, and at its closest neighbors. Closest neighbor is defined on the basis of the frequency that two semantic types are expressed by one and the same construction in the entire subordination database: given semantic type X (e.g. temporal relations) and the set of constructions Y that can encode this type in the entire database, what is the most frequently occurring other semantic type that is expressed by the set of constructions Y? Given these frequencies we can expand to include the closest neighbor(s), and take semantic closeness as a parameter into the equation.

In Table 12.5, an index of semantic closeness is presented in the form of an absolute number of shared constructions per semantic type (see Table 12.1 above) for the entire database. For each semantic type, the two closest neighbors are highlighted in different shades of grey. Table 12.5 shows particularly strong connections between, one the one hand, the relative constructions and, on the other, temporal/reason/condition constructions, and to a lesser extent also with constructions of purpose relations. For complementation strategies the bonds between phasals, modals, and desideratives seem rather strong, as well as those between knowledge, perception, and again

desideratives. If there are traces of contact to be found, we particularly expect them between these three types of semantic clusters.

These different groupings can in turn be correlated to different morphosyntactic forms of the constructions. In particular, I will look at the following parameters:

1. the type of nominalization (participant versus event nominalizations);
2. the expression of core arguments as possessors;
3. case marking.

The subsections are organized according to these three formal parameters, in the order given above, followed by a final section that discusses other issues to do with nominalization.

4.1 *The type of nominalization*

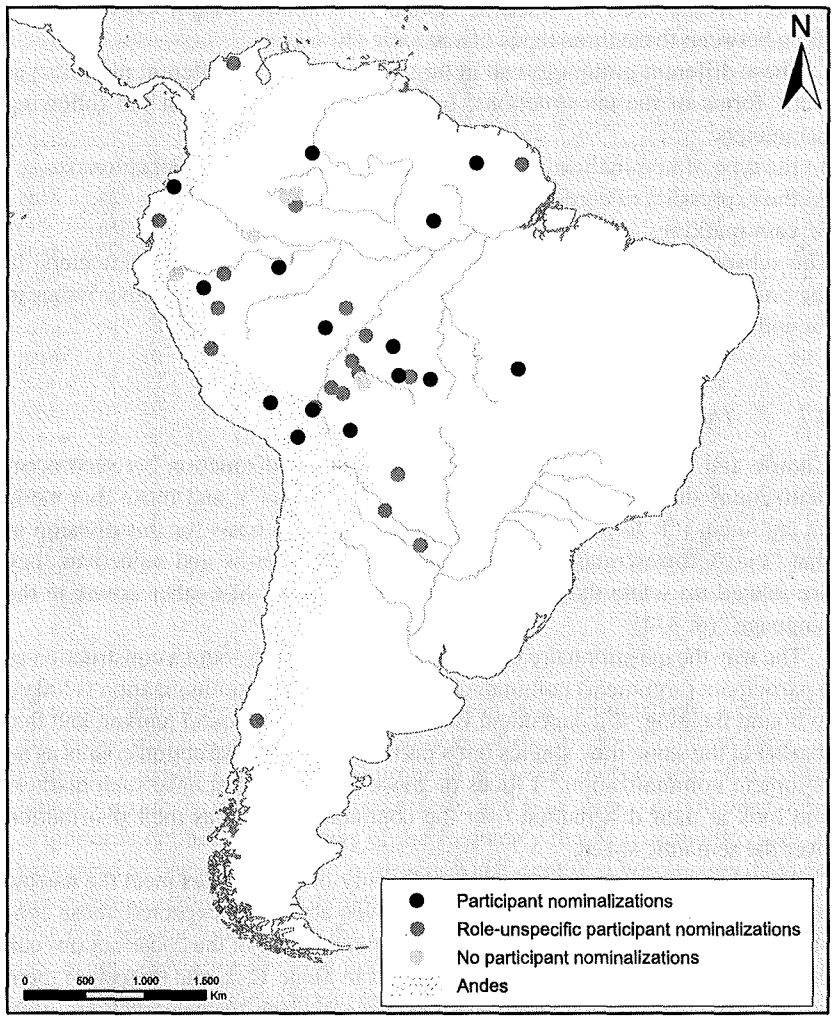
Comrie and Thompson (2007: 334) make a major distinction between nominalizations that name an activity or state ("A forms"), and those that name an argument ("B forms"). They furthermore claim a basis for this division in that "the A forms retain certain properties of the verbs and adjectives they are related to, while those in B behave syntactically like other nouns in the language" (p. 334).

The way the questionnaire is set up, whether or not to count a construction as a participant (argument) nominalization or as an event nominalization is linked to bound flagging. If a dependent EDU is marked by a bound marker, and that marker at the same time singles out a participant, the construction counts as an argument nominalization.[7] I focus on those languages that have such markers and look at their distribution over the continent, as well as their distribution over the semantic space.

A total of thirty-one constructions in twenty-one languages meet the narrow definition of participant nominalizations given above. As expected, these constructions are highly skewed in terms of semantics. All of the constructions can encode relative relations, one of the clusters in Table 12.2, and sixteen of them are exclusively used for relative relations. Nevertheless, the constructions differ in terms of the other semantic relation types they can encode, with purposive, spatial, and temporal relations as the most common non-relative semantic types.

There are two broad strategies that participant-nominalizing languages follow in the relativization of core arguments: (i) the underspecification of participant-denoting nominalizers, and (ii) the use of a paradigm of role-specific

[7] This definition is rather narrow and ignores, for instance, unmarked nominalizations or nominalizations marked by a free marker, and it is restricted to core arguments. However, it captures the most common patterns found in the corpus, and can therefore be expected to give meaningful patterns.



Map 12.1 The use of participant nominalization as a relativization strategy

nominalizers, specifying the semantic role of the relativized argument in the relative clause. The three groups are indicated on Map 12.1: white for no participant nominalizations as relativization strategy; black for those languages that do have participant nominalizations; and grey dots for those languages that have participant nominalizations in a construction where there is a semantically non-specific derivation.

To illustrate this latter difference, consider examples from Desano (Tucanoan) and Kamaiurá (Tupí-Guaraní), which represent the two types. In Desano, there are animate and inanimate nominalizers. Normally, the animate nominalizers yield an agentive relativization and the inanimates a patientive one, but this is not necessarily so, and as a consequence, animate patients yield ambiguous nominalizations (Miller 1999: 142):

(1) bu?e-gɨ
study-NLZ.M.SG
'the one who teaches/the one who studies'

In Kamaiurá, on the other hand, there are different nominalizers depending on the role of the relativized argument in the relative state of affairs. There are separate markers for deriving S (*-ama'e*), A (*-tat*), and P (*-ipyt*) arguments. Example (2) illustrates the S argument nominalizer (Seki 2000: 179).

(2) a-mo-y'u rak akwama'e-a i-'ywej-ama'e-her-a
1SG-CAU-drink at man-NUC 3-be.thirsty-NLZ-PST-NUC
'I made the man who was thirsty drink.'

As mentioned above, some languages allow for other semantic relations to be expressed by these participant nominalizations. These extensions basically follow along the same lines as those mentioned above: non-specific versus paradigmatically opposed specific markers. An example of the first type is the suffix *-taĩ* in the Jivaroan language Aguaruna, which singles out a participant, broadly defined as non-S/A. The precise interpretation depends on whether it carries a case marker or not (Overall 2007: 435).

(3) a. [buukɨa paka-taĩ-numa] ɨhɨų̃ a-u
[skull peel-**NON.A/S:NR**-LOC] CAUS+arrive-REL
'He brought them to the place where skulls were skinned (to make shrunken heads).' (6:3:32)

b. [iwa wampatʃi aintsu iŋkɨpa-taĩ-utʃi-hĩ]
[*Iwa* backpack person put-**NON.A/S:NR**-DIM-PERT:1PL/3]
'his backpack that Iwa puts people in'

An example of the second type is the Nambikwaran language Mamainde (Eberhard 2009: 523–524), where different classifiers, which have a derivational function, can mark different roles in the state of affairs.

(4) a. Paulo-so?ka **wanih-so?ka** kajauka hai?ka wanũn set-so?ka
paulo-NCL.HUM **tell-NCL.HUM** white.man language good speak-NCL.HUM
ka?jãin?-ø-tʰunna-wa
write-S3-FUT2-DECL
'Paulo, the teacher, the one who speaks the white man's language well, he will write.'

b. **anu?ka-hen-ã** eu-khit-ten-lat.a-O-wa
gather-NCL.TIME-FNS see-S1.PL-DES-S3-PRS-DECL
'When we gather together, we will see (about that).'

There are a few potential areal patterns on Map 12.1: (a) the south-central and north-central Andes and foothills (Cuzco Quechua, Huallaga Quechua, Aymara, and foothill languages Leko and Yurakaré and in the north-central area Awa Pit, Aguaruna, and Huallaga Quechua), (b) Rondônia and adjacent areas in eastern Bolivia (Baure, Itonama, Mekens, Mamaindê, Karo, and Apurinã), and (c) the border area between Colombia and Brazil and northeastern Peru (Puinave, Tariana, Desano, Miraña, Urarina). All three of these loosely defined areas are associated with linguistic areas and diffusion of linguistic features, in respective order: the Andean area (see e.g. Torero), the Guaporé-Mamoré area (Crevels and Van der Voort 2008),[8] and the Vaupés (Aikhenvald 2002). In particular, there seems to be an Andean tendency for agent nominalizations that can be used as relative clauses, but specific participant nominalizations also occur throughout the Amazon. Semantically neutral markers or strategies are found in some adjacent languages (Itonama and Baure in northeast Bolivia; Desano and Tariana in the Vaupés area in the border area between Colombia and Brazil; and Miraña slightly further off, in the border area between Colombia and Peru).

Furthermore, a functional equivalence between participant nominalizations and relative clauses seems to be a genealogical trait of a few large families, such as Quechuan, Aymaran, Tupian, and Cariban. The general picture, therefore, seems to be a mix of the fact that some of the most widely dispersed families have this characteristic, and that the trait may also have spread through contact in a number of more regional environments.

A curious final point for this section is the fact that there are four languages, spoken in non-adjacent areas, that permit the participant nominalization to mark same-subject purpose clauses. These constructions are cross-linguistically not very common. The examples come from Cuzco Quechua (Lefebvre and Muysken 1988: 22), Desano (Miller 1999: 153), and Kamaiurá (Seki 2000: 188), respectively.[9]

(5) a. mikhu-q hamu-ni (Cuzco Quechua)
eat-AG come-1
'I come to eat.'

[8] The extent of this area, especially towards the west in Bolivia, is unclear – and is argued to also include the foothill languages – but the clearest areal patterns seem to be found in Rondônia (see Muysken et al. in press).

[9] The fourth language is Huallaga Quechua which, since it is related to Cuzco Quechua, is not represented in the examples.

b. wai wẽhẽ-rã wa-rã ba-bo-rã (Desano)
fish kill-AN.NLZ.PL go-HORT.IMP eat-POT-AN.NLZ.PL
'Let's go kill fish in order to eat!'

c. morerekwar-a je=r-enõj je=r-etsak-ar-am (Kamaiurá)
chief-NUC 1SG=REL-call 1SG=REL-see-NLZ-ATTR
'The chief called me to see me.'

4.2 Possession

An alternative way to express participants in nominalized constructions is by encoding them as possessors. Typological research suggests that S, A, and P participants are all potentially expressed as possessors in nominalizations, but that subject possessors (S/A) are more likely and more frequent than object possessors.

One particular difficulty that arises for South America is that possessors are formally often expressed in the same way as one of the core arguments. For the coding of the questionnaire, this means that there are three answer categories for both subjects and objects: either they are not expressed as a possessor, they are expressed as a possessor, or it is impossible to tell because there is no formal difference between the expression of a possessor and a subject/object.

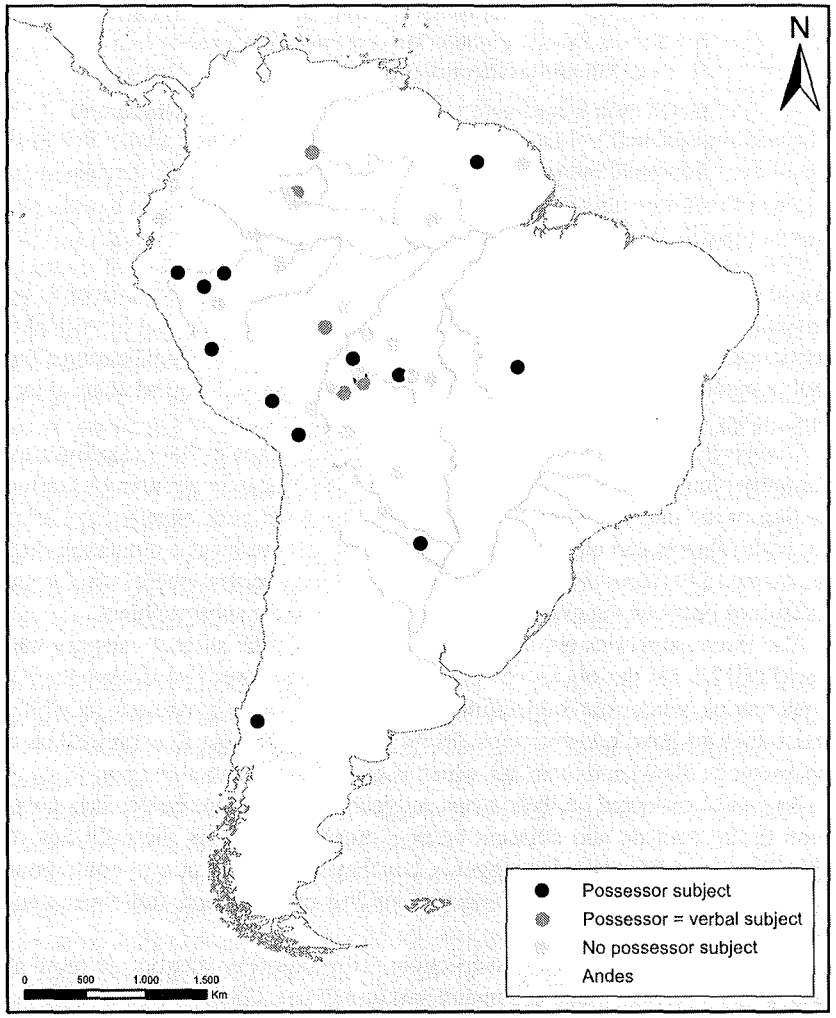
The three categories are shown on Map 12.2 for the subject category and on Map 12.3 for the object category, with the languages that do not have a construction where the subject/object is expressed as a possessor in white, those that do have constructions where the subject/object is expressed as a possessor in black, and those for which it is impossible to tell in grey.

As can be observed on these maps, subject possessors are particularly common in the Andean and adjacent areas – presumably under the influence of Quechuan and Aymaran languages – but they also occur in non-contiguous spots in the Amazon. Object possessors are less common and, moreover, geographically more scattered.

In terms of semantics, the constructions with subject possessors are more or less divided over the range of semantic relation types; the most frequent type is object relativization, illustrated by the contrastive pair from the isolate language Itonama (Crevels 2010: 688), where the b-example is a relative clause, with the subject expressed in the same way as a possessor.

(6) a. k'i-chuduwa'-na lauro chamaye
APPL-buy-NEUT Lauro manioc
'Lauro bought manioc.'

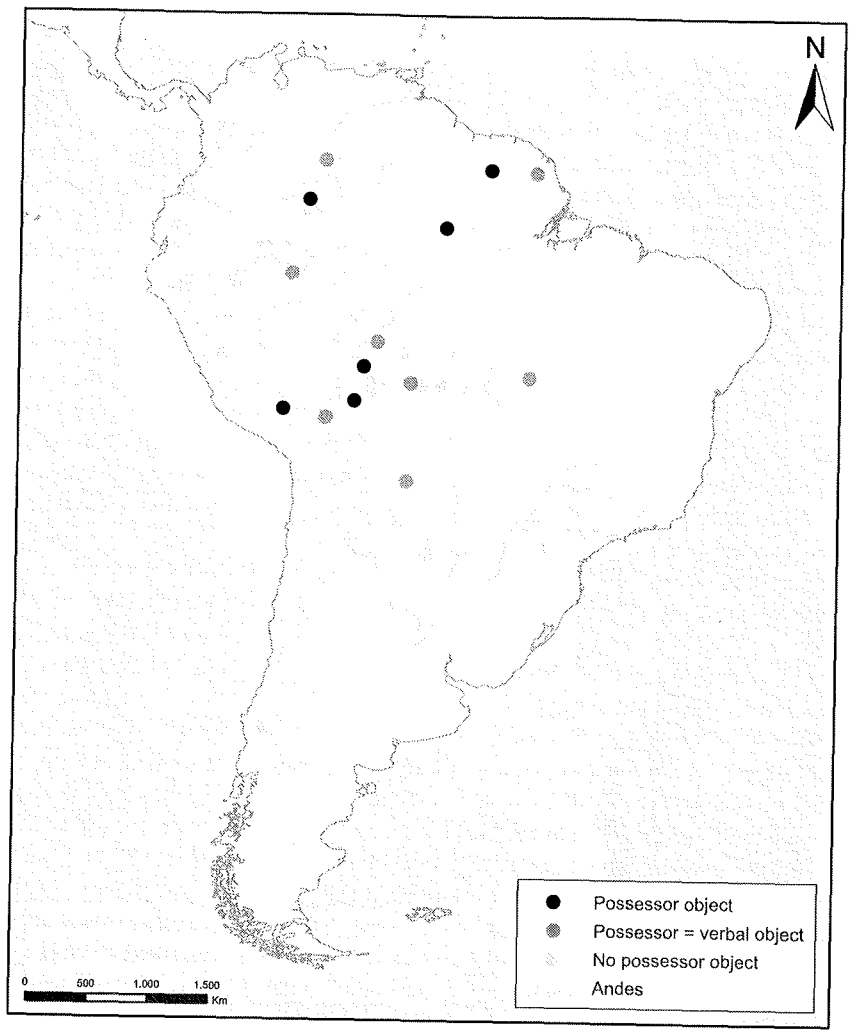
b. lowo'-tya chamaye ah-mi-k'i-chuduwa'-te lauro
be.rotten-STAT manioc 3-REL-APPL-buy-CNT Lauro
'The manioc that Lauro bought was rotten.'



Map 12.2 The encoding of notional subjects as possessors in subordinate clauses

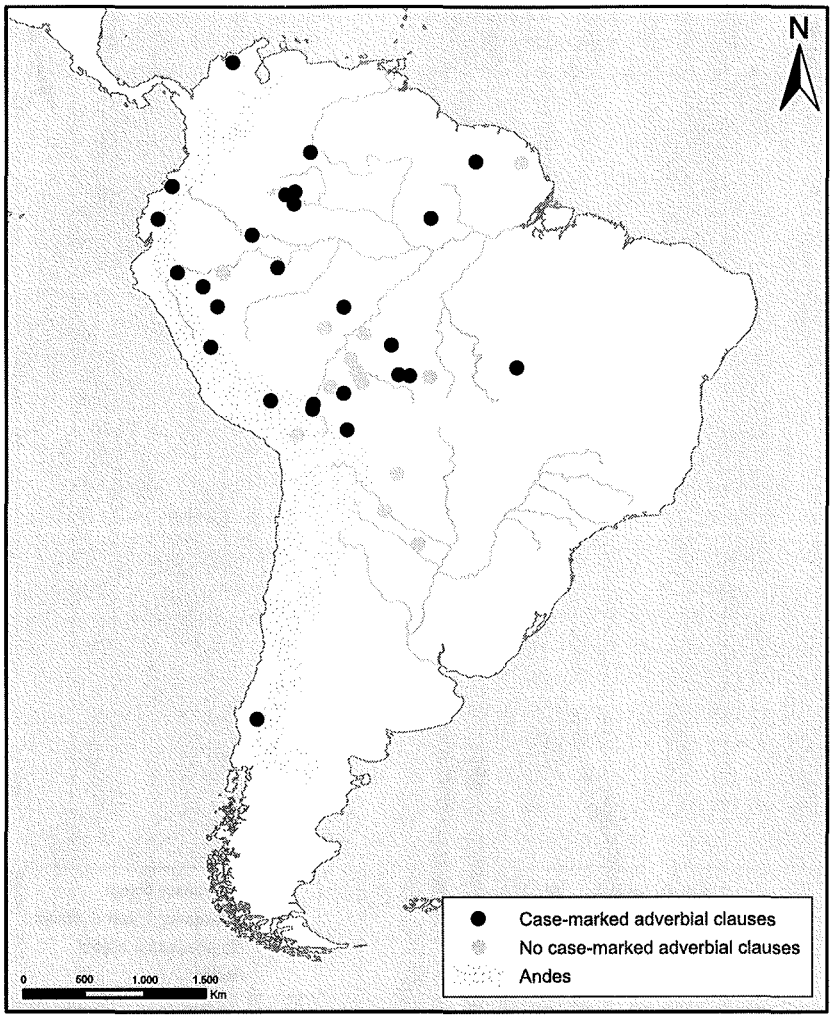
Other slightly more frequent relation types are temporal, reason, purpose, and desiderative relations, partially following Cristofaro's case hierarchy given above.

Given their infrequent occurrence, not much can be said about the semantics of constructions with object possessors. Moreover, the few constructions are more or less evenly divided over the semantic types. Both Cariban languages



Map 12.3 The encoding of notional objects as possessors in subordinate clauses

in the sample, Tiriyó (Meira 1999) and Hixkaryana (Derbyshire 1979), have constructions with object possessors. This can be connected to a more general characteristic of nominalizations in Cariban languages which follow an ergative pattern in the sense that it is the absolutive argument that is expressed as a possessor (Gildea 1992: 125).



Map 12.4 The use of case marking to form adverbial clauses

As a general conclusion of this section, it seems that expressing core arguments as possessors is possibly areally diffused in the case of the Andean area, with the Quechuan and Aymaran languages as the most likely agents of the spread. Object possessors are rarer, and more scattered geographically, but Cariban languages in general seem to have absolutive possessors in their nominalized clauses.

4.3 Case and adpositions

One of the more common nominal features acquired by nominalized predicates is the ability to take case markers, or to be the object of an adposition. In fact, all languages of the sample that have case markers and/or adpositions use these in the formation of complex sentences, with the possible exception of Tapiete. It is therefore not very insightful to project this onto a map, so rather than that, I have chosen to look at oblique case only, used in the formation of adverbial clauses, as shown in Map 12.4.

As can be seen, the majority of languages can form adverbial clauses with case markers or adpositions. This makes this type of construction particularly interesting from the perspective of this chapter, as it is a potential candidate for diffusion. Table 12.6 takes a closer look at the case/adposition-marked adverbial clauses in the sample, with each column indicating a different type of adverbial relation and, for each language, the case marker(s) or adposition(s) that can be used to form the respective adverbial relation type. Empty cells do not necessarily mean that cases or adpositions are not used to express those relation types but can also indicate a lack of information. The table describes the potential of constructions to take oblique case markers, not the obligatoriness of the markers. Furthermore, the information only concerns the semantic relation types that are considered in the questionnaire.

As shown in Table 12.6, temporal, reason, and locative clauses (partly corresponding to the "adverbial" cluster in Table 12.5) in particular tend to be marked with an adposition or a case marker. An often observed strategy is the extension of locational markers to encode temporal relations. Some of the languages that follow this strategy are spoken relatively close to each other (Hup and Tariana in the Vaupés area; Huallaga Quechua and Shipibo in northeastern Peru; Cuzco Quechua, Mosetén, Leko, and Yurakaré in the south-central Andean foothills). Others, such as Mekens and Tiriyó, are more isolated geographically. It may be that contact with members of the Quechuan family has promoted the spread of spatial markers to encode temporal clauses. Another recurring strategy is to use instrument markers for reason relations. The languages that do this, however, are not spoken in a shared vicinity.

In summary, case marking, or the use of adpositions, is a common strategy in South American languages to indicate relationships between events. Some of the sub-structures may be connected to proposed linguistic areas, such as the Andean area and the Vaupés. Again, Quechuan languages may have promoted the spread of this feature.

5 Conclusion

I set out to evaluate the claim that nominalization as a subordination strategy has spread through South America by diffusion through contact, rather than

Table 12.6 *Non-core case markers and adpositions used to form adverbial relations*

	TMP	REA	LOC	PRP	CND
Aguaruna			-numa 'LOC'		
Awa Pit		=akwa 'because'	=ki 'at'		
Cuzco Q	?	-rayku 'cause'	?	-paq 'ben'	
Desano	pɨ?rɨ after kore before bẽrã with		-ge 'loc'		
Hixkaryana		=ke because	=hona 'towards'		
Huallaga Q	-pita abl -chaw loc	-pita abl	-chaw loc -man dir	-paq 'ben'	
Hup	-V́t obl -an dir		-V́t obl -an dir	-V́t obl	
Ika			-ekɨ loc		
Jarawara	jaa 'peripheral'	jaa 'peripheral'	jaa 'peripheral'		jaa 'peripheral'
Kamaiurá		r-ehe 'about, wrt'	-ipe 'loc'	wi 'ablative' (avertive)	
Karo	=kəy dat				
Kwazá	-ko instr	-ko instr			
Leko	-ra loc	-ra loc	-ra loc		-ra loc
Mapudungun		-mew instr	(-mew instr)		
Mekens	=ese loc	=ese loc =eri abl			=ese loc
Miraña		-lií(hye) ben	-vu dir -tu abl		
Moseten	-tom com =ya adess			-dyesi', -dyeti' ben	
Movima	n- obl			n- obl	
Puinave	-a all		-a all	-a all	-a all
Shipibo	pekao after (incl loc)				
Tariana	-se loc -nuku non a/s		-se loc		
Tiriyó	=htao loc	=ke ins			=htao loc
Tsafiki	several locative				
Urarina	baja after bana when				
Yanesha'		-ot loc			
Yurakaré	=jsha abl	=la ins	=y loc =chi dir		

through chance or genealogical inheritance. In order to meet this challenge I tried to answer two questions, repeated here:

(i) Is the distribution of nominalized subordinate clauses geographically skewed towards South America?

(ii) Is there variation within the group of nominalized structures, and is that geographically skewed?

On the global level, question (i) can be answered positively: the occurrence of nominalizations as subordination strategies is significantly higher in South America than would be expected on the basis of Cristofaro's (2003) global sample. This fact alone rules out chance as a possible explanation. Within South America, since almost all languages of the sample have nominalized constructions that can be used as a subordination strategy, there is no clear geographic skewing.

The first part of question (ii) can also be answered positively, as can be seen by only a superficial look at the NeighborNet in Figure 12.1. The second part of question (ii), whether the variation is geographically skewed within South America, is less clear. I reviewed three formal parameters along which nominalizations can differ from each other. In particular, participant nominalizations and case marking are very common strategies. Assuming a diffusion through contact scenario, the widespread occurrence of participant nominalization may be related to a combination of the fact that the major families (Quechuan, Tupian, Arawakan, Cariban) have these structures, and the fact that these features have spread in several smaller areas, such as the Vaupés, the Andean area, and Rondônia (the Guaporé-Mamoré). A similar account can be given for the use of case markers to form adverbial relation types, especially for the Andes. Moreover, the semantic coherence of these groups of constructions makes a spread scenario more plausible. With respect to possession, the semantic coherence is less clear, and the occurrence of core argument possessors is also less pervasive. In particular, agent possession seems common in the Andean area and adjacent zones.

The fact that the Andean area is so dominantly present in all of these areas goes against Dixon and Aikhenvald's (1999: 10) claim that clause nominalization is an Amazonian, and not an Andean, phenomenon. The patterns furthermore only partly confirm Crevels and Van der Voort's claims for subordination through nominalization as an areal feature for the Guaporé-Mamoré area. In the first place, as we have seen, clause nominalization is extremely common, and occurs well beyond the Guaporé-Mamoré area, and second, coherent patterns for the linguistic area itself seem to occur mainly on the Brazilian side of the area.

These patterns do not give us a definitive or direct explanation of the skewed distribution, but they are consistent with a scenario of spread through contact: not as the result of a continent-wide spread region, but rather as the result of

several smaller spread zones, and through a few language families with major extensions, like Quechuan, Tupian, and Cariban. The patterns found do not completely rule out an inheritance-based account, but because nominalized structures are found throughout the continent and across language families and stocks this would mean that the predominance of nominalizations is an extremely old pattern, and the variation found within the group of nominalized structures does not suggest extreme stability for this structure. Another possible reason for the predominance of nominalized clauses is that it is in a dependency relation to some other widespread, more fundamental structural feature of South American languages. This question falls outside the scope of this chapter, and is left for further research. Further research should also make clear whether similar patterns of regional spread can be found for the underrepresented areas in the sample, in particular towards the east.